Chapter 2
Utilizing "Omics" Tools to Study the Complex Gut Ecosystem

Anthony Fodor

Abstract In a healthy gut, the immune system tolerates a diverse microbial commensal community avoiding inappropriate inflammation responses and minimizing the presence of pathogens. When the balance between host and microbes is disrupted, risk for disease increases. There is mounting evidence that microbial dysbiosis is a substantial risk factor for common gut diseases including IBS, IBD and colorectal cancer. Understanding this dysbiosis is challenging because of the extraordinary complexity of the gut ecosystem and the tremendous variability between healthy individuals in the taxa that make up the human microbiome. Advances in technology, especially sequencing technology, are beginning to allow for a full description of this complexity. In this review, we consider how new "omics" technology can be applied to the study of the gut ecosystem in human and animal models with special consideration given to factors that should be considered in the design of experiments and clinical trials.

Abbreviations

ATP	Adenosine triphosphate
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
FDR	False Discovery Rate
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome

A. Fodor (✉)
Bioinformatics and Genomics, UNC Charlotte, 9201 University City BLVD, Charlotte, NC 28223, USA
e-mail: afodor@uncc.edu

M. Lyte and J.F. Cryan (eds.), *Microbial Endocrinology: The Microbiota-Gut-Brain Axis in Health and Disease*, Advances in Experimental Medicine and Biology 817,
DOI 10.1007/978-1-4939-0897-4_2, © Springer New York 2014

OUT	Operational Taxonomic Units
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid

Introduction

In the healthy gut, host immune processes tolerate a diverse commensal population avoiding excessive inflammation responses and minimizing the presence of pathogens. However, there is compelling evidence that microbial dysbiosis—an imbalance in the microbial community—plays a formative role in many diseases of the gut including IBD [1], IBS [2, 3] and colorectal cancer [4, 5]. All gut microbes are acquired from the external environment and for the first 3 years of life the diversity and complexity of the gut microbial community steadily increases [6]. By the 3rd year of life, the microbial community is more stable but numerous studies have repeatedly shown that there is a high degree of individual variation in the microbial community between different people [7–13]. The factors that determine why different people end up with such different microbial communities are poorly understood, although twin studies suggest that host genetics does not exert substantial control over the composition of the microbial community [14].

If we are to understand how host and microbes together produce the full spectrum of health and disease phenotypes, we will need to determine which alleles are represented and expressed in the host, which microbes are present and where in the gut microenvironment the microbes are found and, for both host and microbes, how genes are expressed to produce metabolites within activated pathways. To understand the state of the human and microbial ecosystem in the gut, therefore, requires an accounting of an ecosystem of phenomenal complexity. There are on the order of three billion base pairs in the human genome [15], but there are ~10 times more bacterial cells within the human body than bacterial cells [16] and encoded within the genomes of those microbial cells is likely more than 100 times more distinct genes than are encoded within the human genome [17]. And, of course, only knowing the genome sequence of either host or microbes by itself does not tell us which genes are expressed or where or when or how epigenetic changes to genomes influence pathway structure and function. Within the last decade, there has been explosive growth in "omics" technologies that are allowing us to begin to approach an initial accounting of this tremendous complexity. Development of these technologies have primarily, but not exclusively, been driven by the stunning drop in the cost of DNA sequencing. Only 10 years ago, the cost of sequencing a megabase of DNA was well over $1,000. Today, it is less than $0.10 and there is every reason to think that this greater than exponential drop in cost of sequencing will continue into the future (http://www.genome.gov/sequencingcosts/). Newly armed with ever more affordable sequencing technology, biologists have begun to characterize in detail the complex microbial gut environment. In this

review, we will discuss the technologies that are making this exploration possible together with the experimental and bioinformatics challenges inherent to performing studies that try to link the state of the microbial community to host disease phenotypes.

16S Sequencing Is an Economical Way to Ask “Who Is There” for Both Common and Rare Taxa

For nearly 30 years [18], microbial ecologists have been using sequencing of the 16S rRNA gene to ask which microbes are present in complex microbial environments. The 16S rRNA gene is among the most conserved genes in bacterial genomes. It is especially useful for phylogenetic characterization because it consists of a number of “variable regions”, which tend to be different in different bacteria, separated by “conserved regions”, which tend to be the same across a wide phylogenetic space. The conserved regions can be used to place PCR primers that sequence across the variable regions, yielding a surprisingly informative degree of phylogenetic information from minimal sequencing effort. Before the advent of next-generation sequencing, capillary-based Sanger sequencing was often performed on clone-libraries created from the 16S gene. With a read length on the order of 1,000 basepairs, a paired-end Sanger sequencing strategy could sequence the entire 16S rRNA gene. This approach has been widely utilized and successfully generated descriptions of microbial communities both associated with the human microbiome [19, 20] and external environmental microbial communities such as soil and ocean.

Despite these successes, the cloning approach suffers from several limitations. Because sequences generated from clone libraries are relatively difficult and expensive to generate, studies that characterized microbial communities via sequencing of clone libraries generally could only achieve on the order of 100 16S sequences per sample, and only then with a great deal of expense and effort. Next generation sequencing eliminated the need for the laborious cloning step even as it offered nucleotide base costs that were orders of magnitude cheaper than Sanger sequencing. Next generation sequencing platforms exploit massively parallel chemistry in which numerous sequencing reactions are run at the same time and the results captured with a computer camera. Because many sequencing reactions are run in parallel, next generation sequencing platforms such as Illumina and 454 generate sequences much more quickly than older dye-termination based technologies. In 2005, the year in which the 454 sequencing platform was described in a Nature paper [21], there were ~136,000 16S sequences cataloged in the Ribosomal Database Project (http://rdp.cme.msu.edu/download/posters/ASM2005.pdf). Today, using the Illumina HiSeq platform, we can routinely generate 100 million 16S sequences for a cost of only a few thousand dollars [4, 22, 23].

This ability to generate with next generation sequencing in a single experiment more sequences than had been accumulated world-wide in decades of dye

termination sequencing provides an enormous opportunity to interrogate complex ecosystems, such as the human gut, while maintaining a sensitivity to detect even rare taxa. But it brings with it significant bioinformatics challenges. Some of these challenges involve finding the hard-disk space and network capacity to handle these large volumes of sequence data. Without proper planning for these mundane considerations, it is not uncommon for the initial analysis of metagenomics projects to be severely impacted. There has been considerable recent interest in developing cloud computing capacity to handle these challenges [24] and investigators considering generation of large sequence datasets may wish to explore storing and analyzing their data in the cloud [25].

Bioinformatics challenges can also arise from the short read length inherent to the currently popular next-generation platforms. The early 454 platforms had a read length of only ~100 basepairs [26] and the initial 454 pyrosequencing characterizations of ocean microbial communities therefore utilized this read-length [27, 28]. Recent Illumina platforms, while many orders of magnitude cheaper than 454 sequencing, also have a read length of only 100 basepairs, but 16S sequences of this length can clearly distinguish the microbial community in inflamed and non-inflamed mammalian guts [4] showing the utility of even such short reads. Bioinformatics simulation studies have shown that the information that is available in short reads can be reasonably close to the information available in full length sequences [29], with the V1–V3 and V3–V5 regions of the 16S rRNA considered to be especially appropriate targets given read-lengths of a few hundred base pairs [30, 31] such as are now achievable on 454 and Illumina platforms [26].

The ability to use 16S rRNA sequencing to characterize in-depth the microbial community from cohorts of interest allows for the intersection of phylogeny and traditional hypothesis testing in ways that can yield interesting insights into how the microbial community might impact disease. As an example, a recent study used 454 sequencing of 16S rRNA amplicons to compare the microbial community in punch biopsies taken from 33 subjects with colorectal adenomas and 38 control subjects [12]. In total, slightly more than half a million 16S rRNA sequences with an average length of just over 300 basepairs were generated from these 71 subjects. In order to place the information in these sequences into a phylogenetic context, we can build a tree that shows the relationship of the sequences to one another (Fig. 2.1). Each node of the tree represents a cluster of sequences that have on average 97 % identity to one another. Nodes of the tree that are close to one another have sequences that are more similar while nodes that are further from each other. As we would expect, most of the bacteria that we see in the human gut can be assigned to the phyla Bacteroidetes or Firmicutes, although other phyla, notably Proteobacteria which harbors many known pathogens, are also present. For each taxa in the tree, we can form a null hypothesis that the relative abundance of that taxa is not different in the case and control subjects. *P*-values can be generated for each null hypothesis using a non-parametric Wilcoxon test. In setting thresholds for significance, we must be careful to correct for testing multiple hypotheses. Rather than using a simple-threshold of $p < 0.05$, we instead set a threshold based on a 10 % false discovery rate (FDR), where we expect 10 % of the taxa that we call

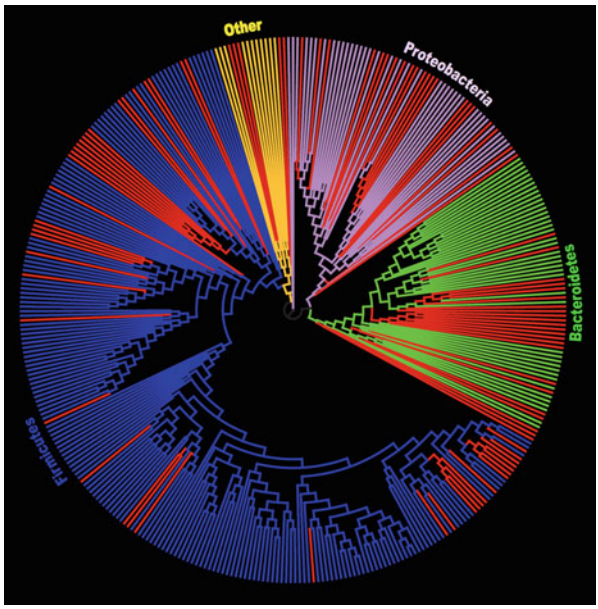

Fig. 2.1 Phylogenetic tree generated from Operational Taxonomic Units (OTUs) representing clusters of sequences with an average 97 % identity from a study of colorectal adenomas in humans [12]. Branches that are colored *red* represent taxa that are significantly different between cases and controls at a 10 % False Discovery Rate (see [12] for methodological details)

significantly different to be false positives. Each taxa that was found to be significantly different between case (adenomas) and control at this threshold is colored red in Fig. 2.1. We see that many of the taxa that are different between case and control are in the phyla Proteobacteria. By creating a visualization that merges phylogeny with canonical hypothesis testing, we are therefore able to begin to implicate specific groups of taxa in disease (see [12] for more information).

Given that early 16S sequencing experiments based on clone libraries could be performed generating less than a hundred reads per sample, it may seem foolish to plan 16S experiments with read depths of over a million sequences per sample. But a simple thought experiment shows that such sequence depths are not inappropriate. Consider *E. coli*, which is a possible driver of colorectal cancer in mouse and human studies [4] but in fecal samples can represent less than 1 % of all sequences collected. On average 100 sequences must be obtained to observe 1 sequence that represents such a rare taxon. If one wishes to study a population with a 1,000-fold range in such a taxon, one must utilize an additional 1,000-fold sequencing depth in order to maintain the full dynamic, quantitative range of sensitivity across people with different relative abundances of the taxon. Finally, in utilizing 454 and Illumina sequences, a barcode method is used in which many samples are put together on the same sequencing run [32, 33]. This procedure can easily introduce a tenfold variation in how many sequences are collected per sample. Putting this together—two orders of magnitude to detect a taxa at average 1 % abundance times three orders of magnitude variation in that taxa between people of different phenotypes times one order of magnitude technical variation in the number of sequences collected per sample—we see that it is not unreasonable to produce and analyze one million 16S sequences per sample.

As technology continues to develop, both read length and read depth will improve allowing for more information to be generated from each sample but also increasing the challenges associated with managing and interpreting so much data. Besides taxonomical considerations, there are many other challenges to the analysis of 16S sequence data including quality assurance steps [34], choosing appropriate clustering algorithms [35] and chimera detection [36]. The setting up of analysis pipelines for 16S sequences has been reviewed elsewhere [37].

Individual Variation Is a Primary Challenge for Studies in the Gut Micobiome

While there are millions of SNP variations between any two non-twin individuals, human genomes have many essential common features that all healthy individuals must share. Every person has to have a working copy of an actin gene, for example, or survival will be impossible. By contrast, the structure of the microbiome does not appear to be essential in the same way. As we will discuss below, mice can be raised in a sterile environment with no gut microbes whatsoever, and while these mice have a great range of phenotypic differences from control mice, they are able to survive [38]. The mammalian gut, therefore, appears to have a certain amount of flexibility with regards to the microbiome. This may explain why, at least at the taxa level as measured by 16S rRNA, a high degree of variability is tolerated in the human microbiome. Perhaps the most dramatic example of microbiome variation was demonstrated by the Human Microbiome Project, which recruited 242 healthy patients and characterized the microbiome by 16S sequencing at 18 distinct body sites [8]. At all the measured body sites, there were tremendous individual differences in this healthy cohort [7, 10]. Moreover, within this cohort, associations between individual taxa and host phenotypes were generally modest. While there were taxa with reasonably strong associations with ethnicity and, as previously observed [39] vaginal pH, associations with phenotypes such as BMI, gender, temperature and blood pressure were moderate at best [10]. It is currently not well understood to what extent the differences in the microbiome associated with ethnicity are driven by genetic or cultural differences, but the possibility of microbial variability produced by ethnicity should be explicitly considered in recruiting cohorts for and powering clinical studies. In general, the modest correlations between healthy human phenotypic variation and microbiome variation suggest that many non-pathological phenotypes are not directly controlled by which taxa are present in the microbiome.

The complexity, individual variation and weak association with phenotypes of the healthy human microbiome represent a substantial challenge for studies that hope to link the state of the microbiome to human phenotypes. If we each have our own unique relationship to the microbiome that defines our own individualized healthy or dysbiotic state, then cross-sectional studies that look across people will

have substantial difficulty in coming to any consistent conclusion. One intriguing idea that has been proposed as a framework to deal with this complexity is enterotypes [40, 41]; it has been argued that much of the complexity of the gut microbiome could be summarized by two or three types of categories dominated by distinct taxa. This hypothesis is enormously appealing as clinical studies could dramatically reduce complexity (and hence improve power) by assigning each participant to one of these pre-defined types before attempting to associate the state of the microbial community to disease phenotypes. Unfortunately, subsequent studies have demonstrated that the presence of enterotypes appears to rely on particular methods of analyzing 16S rRNA data and does not therefore appear to be robust and reproducible in new cohorts [10, 13, 42, 43]. The idea of distinct microbial types likely makes sense for the low-diversity vaginal microbiome [10, 39], but for the more complex gut microenvironment, there appears to be more evidence for a continuum of microbes rather that distinct types.

The variety of gut microbes that will be encountered, and the possibility of only weak associations of taxa with phenotype, must be explicitly considered when powering clinical studies of the human gut microbiome. One approach that may help ease power concerns is to design studies around longitudinal sampling. In a longitudinal sample, each patient in some sense can serve as their own control, which has the potential to reduce variance and hence increase power. Ideally, a longitudinal sampling scheme would recruit a cohort before disease developed and then follow the cohort as some individuals developed disease and others remained healthy. The analysis can then ask both whether the initial state of the microbial community predicted disease and whether changes to the microbial community differ between those who remain healthy and those who develop disease. While this approach is often optimal from the perspective of experimental design, it can be difficult to achieve in practice, especially if the time required to follow a cohort is longer that the length of grant support from funding agencies interested in gut disease.

The Fecal and Mucosal Microbiomes Are Distinct

One great challenge of surveying the gut microbiome, as opposed to more external microbiota such as skin, is that often the microbes that we are most interested in are not the easiest to sample. Fecal samples, obviously, are relatively easy to obtain, but their handling and storage can provide challenging from an operations point of view. Fortunately, it has been demonstrated that issues with how fecal samples are handled, for example how quickly they are frozen, does not appear to have a large effect on the measured 16S community [44]. As an alternative to fecal samples, there has been some interest in utilizing fecal swabs [45], which are easier to collect and store and in the future may be a standard implementation for large clinical studies. No matter how they are collected, however, fecal samples may be inappropriate for studies that evaluate hypotheses regarding the mucosal microbiota.

For example, a recent paper has suggested that microbial DNA may be more present in cancer samples than in non-cancer [46]. Presumably, the microbial invasion that would explain this observation is more likely to occur in the tight contact of host and microbial cells in mucosal material than in the luminal gut. In both human and mouse microbiomes, mucosal and fecal microbiomes generally cluster separately [20] suggesting that there are very distinct luminal and mucosal microbial communities. Obviously, with humans, directly sampling the mucosal microbiota requires an invasive sampling scheme and produces additional IRB requirements, although this collection of internal gut samples can be incorporated into normal colonoscopies. In designing studies, thought should be given to the specific questions being asked and sampling schemes designed accordingly in order to maximize observation of the microbial community most likely to be involved in the phenotype under study.

Mouse Models Have Great Utility but Results Must Be Interpreted with Great Caution

While human association studies are crucial, ultimate evaluation of mechanistic hypotheses about how host-microbe interactions impact disease must be tested in animal models. Because mice can be raised sterile, and then inoculated with a pre-defined microbiome consisting of either cultured [4] or mixed microbial samples [47, 48], gnotobiotic mice allow for testing of hypotheses about how microbes directly cause phenotypes such as cancer [4] or obesity [48]. Despite their power, a number of caveats must be observed when designing and performing mouse microbiome experiments. In particular, once the gavage has been performed, a number of factors not related to the contents of the initial gavage can substantially alter the microbial community. These factors include the cage the mice are housed in [49], the facilities the animals are housed in [50], the amount of time that has elapsed since exposure to microbes [51] and (in animals not raised sterile) the line of maternal transmission [52]. If these factors are not accounted for, they may induce variations in the microbial community that may confound interpretation of experimental design. In a recent study comparing animals gavaged to animals allowed to acquire their microbial community from the environment of the animal facility, it was found that while the initial gavage had an effect on the microbial community, most of the composition of the microbial community was driven by the amount of time that had elapsed since animals were removed from germ-free conditions and the cage in which the animals were kept [53]. Clearly, experimental designs that do not explicitly consider these factors are likely to lead to flawed conclusions and in powering mouse studies, the number of cages, in addition to the number of animals, must be explicitly considered.

Whole-Genome Metagenome Sequencing and RNA-Seq Can Be Used to Interrogate Genome Function

As outlined above, small regions of the 16S rRNA sequence can be surprisingly informative, but there are limits to how much information can be generated by measuring a single gene. The drop in the cost of sequencing has made much more feasible experiments which measure all the genes present in microbial genomes (whole genome metagenome-shotgun sequencing) and experiments which measure microbial transcripts from mixed microbial communities (metagenomic RNA-seq experiments). As is the case for 16S sequencing, initial sequencing effort using Sanger sequencing for whole-genome metagenome experiments required substantial investments of time and expense. An early whole-genome metagenome shotgun sequencing experiment [54] using clone libraries and Sanger sequencing produced ~78 million bases of unique sequence from fecal samples of two human subjects, producing our first look at the genome content of the gut microbiome. Today, through the use of Illumina HiSeq, it is not uncommon to produce ~2 gigabases of sequences per sample, with per sample costs in the hundreds of dollars. As is the case for 16S sequences, therefore, we can now produce in a single experiment more sequences than were produced by multiple labs over years of experiments using Sanger sequencing.

To be of any utility, whole-genome metagenome sequencing generally requires many more sequences per sample than 16S sequencing. This translates both into more expense and a more difficult analysis path. Not only does hard-disk and network capacity need to be found for the large numbers of sequences that will be generated by these methods, but the mapping of individual reads to reference gene databases can require substantial computational times. Investigators wishing to perform whole-genome or RNA-seq on microbial communities must therefore ensure they have adequate computational resources or risk project paralysis in attempting to sift the data once the sequences have been obtained.

Despite the increased overhead and expense of whole-genome sequencing approaches, these experiments can yield great insights into the gut microbial community. An intriguing result from the Human Microbiome Project found that while across body sites and individuals there was great variability in taxonomy (as defined by 16S sequences), if one looks at the fraction of reads assigned to gene functions, they was much more consistency [7]. This result suggests the intriguing hypothesis that while taxa vary substantially in the human microbiome, the gene functions encoded in those taxa are much more constant. Of course, this interpretation of these results is very dependent on the accuracy of functions that are in gene function databases and there has been some question as to how biased these databases may be [55]. Moreover, it is perhaps not surprising that across samples and subjects, the fraction of genes assigned to broad categories such as “ATP synthesis” and “central carbohydrate metabolism” is reasonably constant. It remains an open question how much this high-level consistency is reflected in consistency in specific metabolic pathways. It will be fascinating to watch

resolution of the question as to the best way to biologically interpret gene function annotations as the technologies and approaches that power the study of the human microbiome continue to mature.

If instead of whole-genome sequencing of DNA, RNA is isolated, largely the same informatics pipelines can be used to assign gene functions at the transcript level. Because RNA is much less stable than DNA, these experiments are often more difficult to perform than whole-genome shotgun sequencing, but since message is being measured, rather than just genomic potential for message, the biological insights generated from these experiments can be considerable. In addition to the usual difficulties associated with any RNA preparation, RNA-seq on microbial and metagenomic populations has its own set of challenges. These arise from the fact that unlike eukaryotic mRNA, prokaryotic mRNA does not have a poly-A tail. Message and ribosomal RNA therefore cannot be easily separated by the use of poly-T primers during transcription of cDNA. Strategies that utilize beads that preferentially bind to, or enzymes that preferentially cleave, rRNA have been developed to separate mRNA from rRNA, although these strategies have been found to vary substantially in effectiveness [56]. One strategy that becomes more attractive as sequencing costs drop is to not attempt to separate rRNA from message RNA and simply rely on sequencing depth to characterize the mRNA that may be present in a sample. This strategy has the appeal of simplicity and will also generate a complete rRNA profile, that can itself be useful in taxonomic assignment. Its successful application, however, depends on sequencing being inexpensive enough that sufficient sampling depth can be generated to characterize the small fraction of reads that are message.

For both whole-genome metagenome sequencing and RNA-seq from mixed microbial communities within the human microbiome, there is also the problem of host contamination. The bulk of nucleotides in fecal samples is microbial, but in other tissues the fraction of microbial vs. host DNA and RNA can vary substantially. Again, as sequencing becomes ever cheaper, the strategy of simply applying more sequences and computationally removing human contaminant becomes more attractive, assuming that sufficient computational resources are available to achieve an initial parse of sequence data.

Future Studies Will Integrate Multiple "Omics" Techniques to Generate a Complete Picture of Host and Microbial Pathways

In parallel to the decrease in the cost of nucleotide sequencing, metabolomic and proteomic platforms are continuing to increase in power, robustness and accessibility. In proteomics, a major challenge is identifying spectra and this challenge is only increased in the case of mixed metagenomic communities where the genome sequences that give rise to proteins are not necessarily known [57]. Despite this,

recent efforts have demonstrated not only that proteomics on metagenomics samples is feasible [58] but that the combination of metagenomics and metaproteomics approaches can pinpoint particular host and microbial pathways that are associated with disease [59]. Further integration of these techniques with metabolomics will undoubtedly yield additional insights [60]. The principle challenge of performing these types of studies is the integration of diverse genomics datasets, but this is an area of active research in bioinformatics [61]. We will unquestionably see more and more studies in the future that will combine nucleotide sequencing with proteimic and metabolomic techniques.

While the new world of “omics” and its associated bioinformatics tools are often thought of as the “microscope” through which we can understand the gut ecosystem in all its complexity, the tools of traditional microbiology, having been continuously refined over the last century, are powerful and should not be overlooked. It is often stated that most gut microbes are not cultivable, but a recent study that attempted to systematically cultivate gut microbes from fecal metagenomic samples found that a substantial proportion of microbes that were detectable with 16S sequencing could be cultivated with high-throughput anaerobic techniques [62]. Because these organisms can be introduced into sterile mice, creation of these biobanks of cultivated organisms will allow for explicit testing of hypotheses about which taxa and groups of taxa are associated with disease phenotypes. Moreover, with newly affordable high-throughput sequencing, whole-genome sequences can be easily obtained for these cultivated organisms, which will allow for delineation of which genes and genome regions drive health and disease associations in humans and produce measurable phenotypes in mice. This marriage of classical microbiology with gnotobiotic and sequencing technology will likely prove a powerful tool in the next decade’s attempt to understand how specific pathways are implicated in disease phenotypes.

Conclusion

The gut ecosystem is very complex, but there has been substantial and exciting recent progress in development of genomic and bioinformatics tools that can allow for delineation of that complexity. The initial phase of the Human Microbiome Project focused on utilizing sequencing to characterize variation in healthy adults. As we move into the next phase of the study of the human microbiome, a central focus will be on determining which microbial taxa, genes and pathways are implicated in disease. Careful design of clinical trials and experiments in animal models will be required to overcome the substantial background variation in the gut microbiome and separate confounding variables that are often closely related to the disease categories of interest. A central challenge will be the integration of different types of “omics” data to produce mechanistic descriptions of how host and microbe together produce phenotype.

References

1. Manichanh C, Borruel N, Casellas F, Guarner F (2012) The gut microbiota in IBD. Nat Rev Gastroenterol Hepatol 9:599–608
2. Öhman L, Simrén M (2013) Intestinal microbiota and its role in irritable bowel syndrome (IBS). Curr Gastroenterol Rep 15:1–7
3. Talley NJ, Fodor AA (2011) Bugs, stool, and the irritable bowel syndrome: too much is as bad as too little? Gastroenterology 141:1555–1559
4. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM et al (2012) Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 338:120–123
5. Zhu Q, Gao R, Wu W, Qin H (2013) The role of gut microbiota in the pathogenesis of colorectal cancer. Tumour Biol 34:1285–1300
6. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG et al (2012) Human gut microbiome viewed across age and geography. Nature 486:222–227
7. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214
8. Human Microbiome Project Consortium (2012) A framework for human microbiome research. Nature 486:215–221
9. Spencer MD, Hamp TJ, Reid RW, Fischer LM, Zeisel SH et al (2011) Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. Gastroenterology 140:976–986
10. Huse SM, Ye Y, Zhou Y, Fodor AA (2012) A core human microbiome as viewed through 16S rRNA sequence clusters. PLoS One 7:e34242
11. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. Nature 489:220–230
12. Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L et al (2012) Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. ISME J 6:1858–1868
13. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM et al (2012) Gut microbiota composition correlates with diet and health in the elderly. Nature 488:178–184
14. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A et al (2009) A core gut microbiome in obese and lean twins. Nature 457:480–484
15. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ et al (2001) The sequence of the human genome. Science 291:1304–1351
16. Savage DC (1977) Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31:107–133
17. Yang X, Xie L, Li Y, Wei C (2009) More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. PLoS One 4:e6074
18. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML et al (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci 82:6955–6959
19. Hayashi H, Sakamoto M, Benno Y (2002) Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. Microbiol Immunol 46:535–548
20. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L et al (2005) Diversity of the human intestinal microbial flora. Science 308:1635–1638
21. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380
22. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl Environ Microbiol 77:3846–3852
23. Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD et al (2010) Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. PLoS One 5:e15406

24. Schatz MC, Langmead B, Salzberg SL (2010) Cloud computing and the DNA data race. Nat Biotechnol 28:691–693
25. Dai L, Gao X, Guo Y, Xiao J, Zhang Z (2012) Bioinformatics clouds for big data manipulation. Biol Direct 7:43
26. Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769
27. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM et al (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci 103:12115–12120
28. Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR et al (2007) Microbial population structures in the deep marine biosphere. Science 318:97–100
29. Liu Z, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res 36:e120
30. Kim M, Morrison M, Yu Z (2011) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. J Microbiol Methods 84:81–87
31. Hamp TJ, Jones WJ, Fodor AA (2009) Effects of experimental choices and analysis noise on surveys of the "Rare Biosphere". Appl Environ Microbiol 75:3263–3270
32. Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B et al (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res 35:e130
33. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 5:235–237
34. Huse S, Huber J, Morrison H, Sogin M, Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143
35. Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ Microbiol 12:1889–1898
36. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200
37. McCafferty J, Fodor A (2013) Human microbiome analysis via the 16s rRNA gene. In: Bishop OT (ed) Bioinformatics and data analysis in microbiology. Caister Academic Press, Norfolk
38. Bäckhed F, Manchester JK, Semenkovich CF, Gordon JI (2007) Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. Proc Natl Acad Sci 104:979–984
39. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK et al (2010) Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108(Suppl 1):4680–4687
40. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T et al (2011) Enterotypes of the human gut microbiome. Nature 473:174–180
41. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y et al (2011) Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105–108
42. Jeffery IB, Claesson MJ, O'Toole PW, Shanahan F (2012) Categorization of the gut microbiota: enterotypes or gradients? Nat Rev Microbiol 10:591–592
43. Koren O, Knights D, Gonzalez A, Waldron L, Segata N et al (2013) A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. PLoS Comput Biol 9:e1002863
44. Wu G, Lewis J, Hoffmann C, Chen Y-Y, Knight R et al (2010) Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol 10:206
45. Lautenbach E, Harris AD, Perencevich EN, Nachamkin I, Tolomeo P et al (2005) Test characteristics of perirectal and rectal swab compared to stool sample for detection of fluoroquinolone-resistant *Escherichia coli* in the gastrointestinal tract. Antimicrob Agents Chemother 49:798–800
46. Riley DR, Sieber KB, Robinson KM, White JR, Ganesan A et al (2013) Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. PLoS Comput Biol 9:e1003107
47. Faith JJ, McNulty NP, Rey FE, Gordon JI (2011) Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science 333:101–104

48. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER et al (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027–1131
49. Hildebrand F, Nguyen TLA, Brinkman B, Yunta R, Cauwe B et al (2013) Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. Genome Biol 14:R4
50. Ma BW, Bokulich NA, Castillo PA, Kananurak A, Underwood MA et al (2012) Routine habitat change: a source of unrecognized transient alteration of intestinal microbiota in laboratory mice. PLoS One 7:e47416
51. Gillilland MG, Erb-Downward JR, Bassis CM, Shen MC, Toews GB et al (2012) Ecological succession of bacterial communities during conventionalization of germ-free mice. Appl Environ Microbiol 78:2359–2366
52. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I et al (2012) Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. J Exp Med 209:1445–1456
53. McCafferty J, Marcus M, Gharaibeh RZ, Arthur JC, Perez-Chanona E et al (2013) Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. ISME J 7:2116–2125
54. Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ et al (2006) Metagenomic analysis of the human distal gut microbiome. Science 312:1355–1359
55. Schnoes AM, Ream DC, Thorman AW, Babbitt PC, Friedberg I (2013) Biases in the experimental annotations of protein function and their effect on our understanding of protein function space. PLoS Comput Biol 9:e1003063
56. Giannoukos G, Ciulla D, Huang K, Haas B, Izard J et al (2012) Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol 13:r23
57. Cantarel BL, Erickson AR, VerBerkmoes NC, Erickson BK, Carey PA et al (2011) Strategies for metagenomic-guided whole-community proteomics of complex microbial environments. PLoS One 6:e27173
58. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M et al (2008) Shotgun metaproteomics of the human distal gut microbiota. ISME J 3:179–189
59. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF et al (2012) Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PLoS One 7:e49138
60. Turnbaugh PJ, Gordon JI (2008) An invitation to the marriage of metagenomics and metabolomics. Cell 134:708–713
61. Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol 7:198–210
62. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A et al (2011) Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc Natl Acad Sci U S A 108:6252–6257

http://www.springer.com/978-1-4939-0896-7

Microbial Endocrinology: The Microbiota-Gut-Brain Axis in Health and Disease

Lyte, M.; Cryan, J.F. (Eds.)

2014, XVIII, 436 p. 48 illus., 19 illus. in color., Hardcover
ISBN: 978-1-4939-0896-7